2017 International Conference on Sustainable Information Engineering and Technology (SIET)

Business Intelligence for Physical Examination Platform Service Reporting System

Abba Suganda Girsang
Computer Science Department, BINUS Graduate Program-Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480, Email : agirsang@binus.edu

Evans Andita
Computer Science Department, BINUS Graduate Program-Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480, Email : evans200991@gmail.com

Sani Muhamad Isa
Computer Science Department, BINUS Graduate Program-Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480, Email : sani.m.isa@binus.ac.id

Arie Purnama
Computer Science Department, BINUS Graduate Program-Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480, Email : purnamaarie95@gmail.com

Aditya
Computer Science Department, BINUS Graduate Program-Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480, Email : adityatang127@gmail.com

Ferico Samuel
Computer Science Department, BINUS Graduate Program-Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480, Email : ferico55@gmail.com

***Abstract*—Currently, the capability to present reports or data accurately, fast and insightfully is highly required for the company to make data-driven decision. This paper is intended to solve the problem in one company which serves more than sixteen professional companies in South East Asia area. This report is useful to make decisions in this company. The method chosen to develop the data warehouse along with its analytics and reports is Kimball methodology which has been introduced since the mid-1980s and has been used by a lot of prior researchers. As the data can be displayed in various form as it needed, the stakeholder can make the data-driven decision which benefits this company to perform better in the market.**

Keywords—Data warehouse, kimball, report, physical examination

I. INTRODUCTION

The data warehouse is relatively a new field of study in computer science which is central repositories of integrated historical data typically in an enterprise in one single place which is used for reporting and data analysis [1]. Research towards this topic is incredibly popular in last few years which shows high interest in the topic. The rise of data warehouse itself was affected by big data phenomenon nowadays. The term big data itself has been in use since 1990 by John Mashey which refers to data sets with sizes beyond the ability of commonly used software tools to capture, curate, manage, and process data within a tolerable elapsed time [2]. The emergence of data warehouse along with the big data terms has resulted from the enormous amount of data resulted within companies and the abilities of the system to process and store the data as the price of storage are cheaper [3].

This study takes one case study on a company which runs in human resource partner services and serves South East Asia Area since 2004 and is developing new physical examination service for its partners. This kind of service is already common in term of human resource partner company but with years of experience expansion to a new common field is a good option. For now, the company already running the new physical examination service for two countries partnering with several medical checkup providers. Since the company is established for a long time and has its research and development division the company requires a service from which it can extract and view data and reports in a various way. This kind of service is required for the company to make a data-driven decision.

This company currently does not have a reporting system that can generate report data from database in fast, efficient, and automatically. Due to that limitation, the time needed for report creation is relatively long. Currently, all the reporting is made by manually exporting the database to Microsoft excel and analyze further using the same application. Therefore, the implementation of the data warehouse is needed to accommodate precise and easy reporting. Data warehouse and OLAP analysis is completely aligned with the company needs since the resulting analysis is able to display data as various reports as how the company probably need in order to perform data-driven decision making. The term data-decision making itself requires an enormous amount of historical data to be analyzed and watched carefully to support or oppose the current hypothesis and perform decision making based on the analysis made [4].

The methodology used for the data warehouse development is Kimball lifecycle methodology which was conceived since mid of 1980 and first published in the 1990s. This methodology is proven and remains in use by a lot of researchers in the various field [5] [6]. Kimball lifecycle itself has been broadly adopted and have become mainstream industry best practices [7].

978-1-5386-2182-0/17/$31.00 ©2017 IEEE

II. Related Works

The data warehouse is a relational database designed for query and analysis rather than for transaction processing [8]. Data warehouse itself is constructed by integrating data from multiple heterogeneous sources which typically is an online transaction processing (OLTP) database. Data warehouse has the property of being a subject-oriented, integrated, time-variant, and nonvolatile collection of data that used for supporting management decision-making process. Data warehouse which is holding enormous amount of data is stored on an enterprise mainframe server or increasingly, in the cloud nowadays [9].

Data warehouse or also called online analytical processing (OLAP) is market-oriented compared to conventional OLTP. OLAP is very fundamental to making a Decision Support System (DSS) which is any tool used to improve the process of decision making in the complex system. The OLAP has to integrate data from multiple heterogeneous information sources and transform them accordingly into the various multidimensional report which will be used for further analysis and decision making [10].

The process to integrate and bring all of data from various multiple data sources into a single central repository is commonly called ETL (Extraction-Transformation-Loading) process. Within extraction process, data is extracted from different data sources and propagated into data staging area where the data will be transformed and cleaned before being loaded to the final repositories. Data sources, data staging area, and target environments may have completely different data structure formats [11].

As the data warehouse gathers information from multiple data sources, the size of the data that will be stored and processed will be incredibly huge. With remarkable growth in technology and its adoption, the amount of data produced is increasing with every minute. Therefore, the ability to keep hold all of resulted data and scalability of data warehouse in very critical [12].

The scalability itself can be achieved by ensuring the performance both of the query itself and the performance of the server in which the data warehouse is located. With major improvement and shift to the cloud era, many researches have shown that performance of the data warehouse can be increased by utilizing cloud computing. Cloud infrastructure provides flexibility along with performance and scalability to cope with every computation needs of the data warehouse. Data nowadays has become prerequisites for making the decision and driving companies and thus the number of query and analytics request to the data warehouse increased incredibly [13].

Data warehouse in healthcare domain could be more challenging as this domain is always dealing with sensitive data which is governed by privacy rules. The center of transaction processing data scheme is patient but there are situations where provider centric data analytic is needed. Implementation of data warehouse and analytics in healthcare domain can help in improving healthcare quality [14].

For implementing this system, precise reporting for a data-driven decision is needed as the foundation of request for change (RFC). Based on this RFC, comparison between old and new system is done for designing proposed system that meets the needs of the company. After implementation of the new system, evaluation is done to see whether the new system meet the goal of the company [15].

Integrating Information Technology Infrastructure Library (ITIL) with the process of operational, strategic and tactical levels of the company could bring improvements to the dynamic of business. This can be used for decision support, modification for quality improvement based on the collected data. Change management applied to the system to support the modification so it meets the procedure for changing the old process [16][17].

III. Proposed Method

Request for change (RFC) was made by one of the employees of this company, as he looks the needs of the company in future. The main purpose of the RFC is to propose a new reporting system based on data warehouse implementation.

Based on Kimball Methodology which is bottom-up approach, the first step of creating a data warehouse is to do business requirement definition. In this very first step, understanding towards the business process is critical. By understanding the business process and its requirement, we will be able to engage the business, prioritize the efforts and deliver business value. The business requirement definition process is done by conduct interview with business stakeholder to get insight and better understanding of the business itself and further setting goal of the data warehouse implementation.

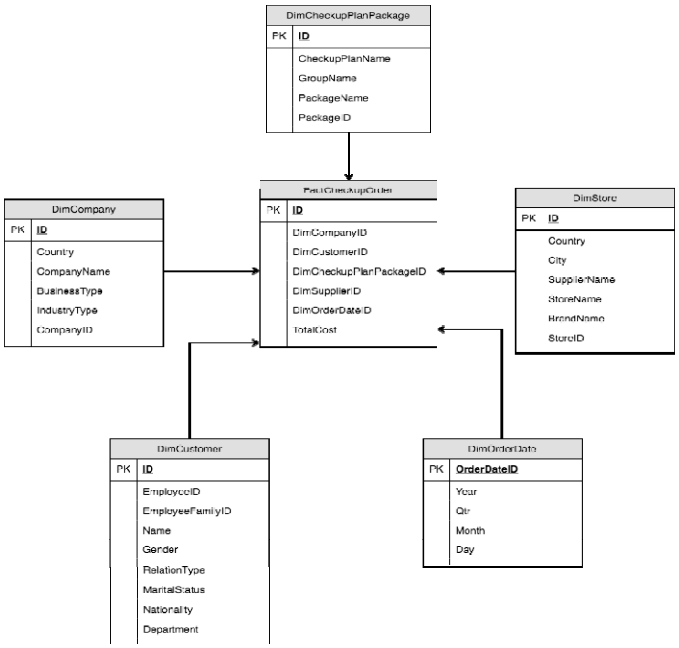

Fig.1 Star Scheme

Based on the understanding, the next step is to design the data warehouse starting from the star scheme. The granularity

of the data is to retrieve total cost spent for each of transaction per customer. After deciding the granularity of the data, dimensions of the data to be analyzed have to be chosen. The result of the dimensions and granularity selection is reflected on the star scheme on Fig.1.

After the star scheme is designed, the next step is to do physical design by creating the database based on the resulted star scheme along with its relationship and fields. The resulted physical database will be used for online analytical processing database. Even it is common to have multiple OLAP databases, in this research, we limit the number of OLAP database to one since the implementation of the data warehouse in this company is new and further comparison and adjustment will be needed.

As we discussed before, some processes are needed to transfer the data from conventional online transaction database to online analytical processing database which are extraction, transformation and loading (ETL). ETL design and development is done next and will result set of queries which the ETL process accordingly. The first process of ETL is to transform the OLTP database into a new database format using the star scheme above. As we can see on Fig.2, the ETL process from OLTP database for employee and employee family table are converted into one customer table.

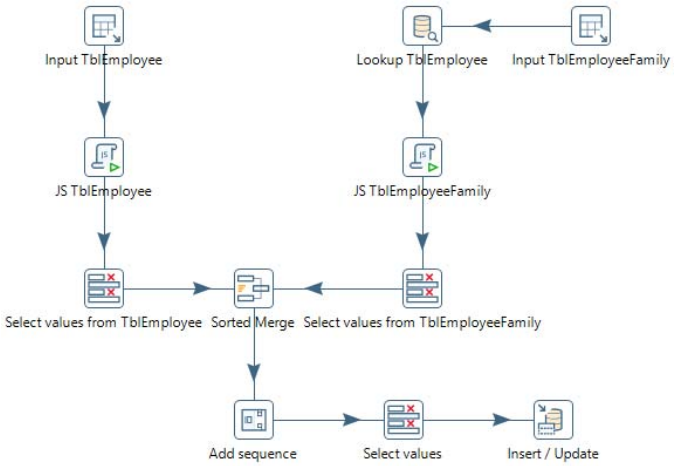

Fig. 2 Loading Customer Dimensions

Then another dimension such as company, store, checkup plan package, and order date are follow to be transformed from OLTP database to OLAP database. All ETL processing and reporting process are using Pentaho Data Integration and Pentaho Businesses Analytics. Fig. 3 is the ETL process of transforming company OLTP data. First, we need to fetch all the OLTP data then we modify the raw data using the JavaScript as seen in Fig 4. The other dimension is performed as shown Fig. 5 and Fig.6. After the modification of raw data, we add database lookup add a sequence to fill in the surrogate key. Last, we transform the data into OLAP table. Fig.7 represents a JavaScript validation used in Fig 3 and Fig 4, this validation is to modify the value of the country code into the country name. In this case, we want to validate two types of country code to their country name respectively.

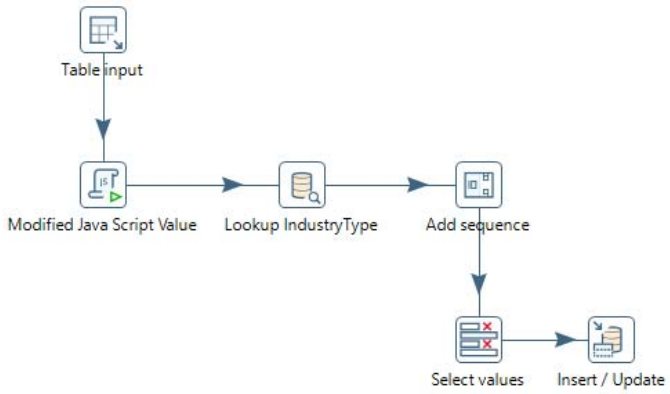

Figure. 3 Loading Company Dimension

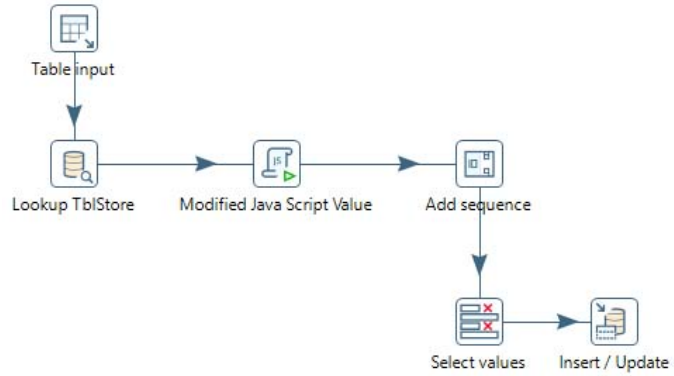

Fig. 4 Loading Store Dimension

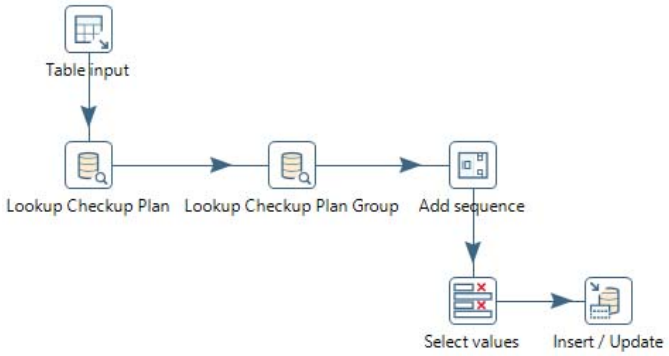

Fig. 5 Loading Checkup Plan Package Dimension

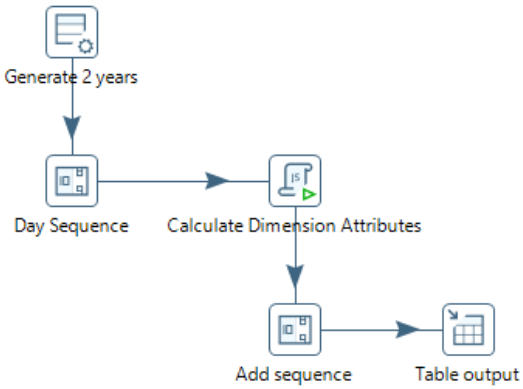

Fig. 6 Load Order Date Dimension

```
var result = (AddressCountryCode == 'CN')
    ? 'China'
    : (AddressCountryCode == 'ID')
        ? 'Indoensia'
        : AddressCountryCode;
```

Fig.7 Company Modified JavaScript Value

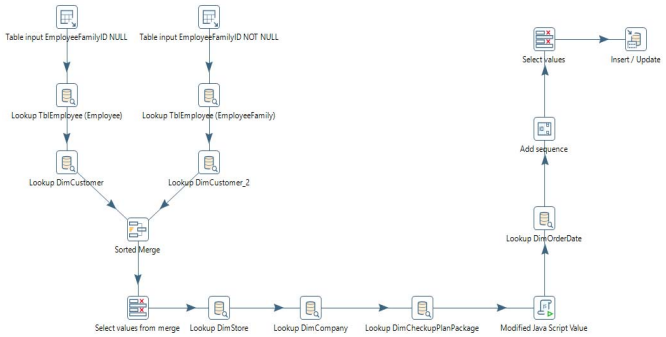

Fig. 8 Loading Fact Checkup Order

The last of ETL process is to fill in the fact table, in doing so we need to find the individual ID in each dimension. Therefore, we need to look up to all dimension table in OLAP database. So then, we can transform the data and convert the data to checkup order fact table as shown Fig.8

IV. Result and Discussion

After the ETL process is done and report dashboard is generated, user can now view the report from the data in various form as the user need. Some sample reports can be viewed on Fig.9 and Fig.10.

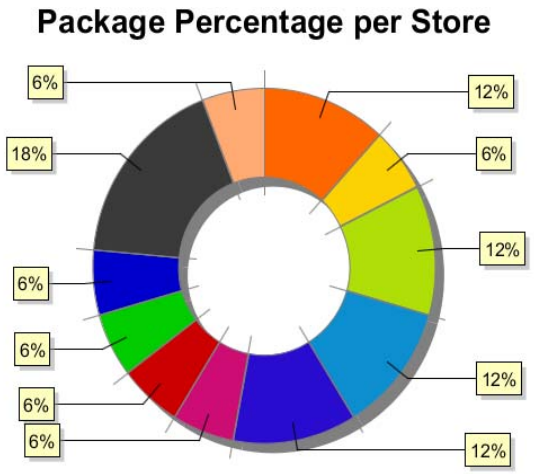

Fig. 9 Package Percentage per Store

Fig. 10 Total Package Over Time

By viewing the report, the stakeholder able to get more insight and therefore able to make data-driven decision making. ETL process will be done twice a day at midnight and midday which considered as real time. Using real-time data supported by various reporting, the analysis resulted will be more reliable and therefore useful to drive the company decision. After the data warehouse system was built, all dimension report can be figured as based on for taking decision. Moreover, it can be built OLAP cube for some dimensions.

After implemented, the old and new system is compared to see the performance as shown Table I. The new system was estimated can make the reporting flow can be done more quickly and more precise.

TABLE I ESTIMATE COMPARE SYSTEM

Matrix	Old System	New System
Time to make the report	More than 2 days	Less than 1 day (near real time)
Ease of use	Manually Processed	Automatically Processed
Data accuracy	Higher human error because data exported and processed manually	More precisely because the data directly connect to the transaction data

V. Conclusion

Based on the comparison result, it is clear that data warehouse implementation for this company boosts the decision-making process by providing more robust, fast, and efficient report. The report itself which can be shown in various form help the stakeholders to get better insight towards the data and do the data-driven decision making. Due lack of time and sample data, there is lack of precise on comparing old system and new system. Advice for next researchers, for implementing data warehouse as a new system in the company, a precise measurement of utility from old system is needed.

References

[1] N. Dedic and C. Stanier, "An Evaluation of the Challenges of Multilingualism in Data Warehouse Development," International Conference on Enterprise Information Systems - ICEIS, vol. 18, 2016.

[2] C. Snijders, U. Matzat and U.-D. Reips, ""Big Data": Big Gaps of Knowledge in the Field of Internet Science," International Journal of Internet Science, pp. 1-5, 2017.

[3] N. Marz and J. Warren, Big Data: Principles and best practices of scalable realtime data systems, Manning, 2015.

[4] J. Kamki, Digital Analytics: Data-Driven Decision Making in Digital World, Notion Press, 2017.

[5] S. N. Murphy, P. Avillach, R. Bellazi, L. Phillips, M. Gabetta, A. Eran, M. T. McDuffie and I. S. Kohane, "Combining clinical and geonomics queris using i2b2-Three methods," PloS one, 2017.

[6] A. Abelló, O. Romero, T. B. Pedersen, R. Berlanga, V. Nebot, M. J. Aramburu and A. Simitsis, "Using semantic web technologies for exploratory OLAP: a survey," IEEE transactions on knowledge and data engineering, vol. 27, no. 2, pp. 571-588, 2015.

[7] R. Kimball and M. Ross, The Kimball Group Reader: Relentlessly Practical Tools for Data Warehousing and Business Intelligence Remastered Collection, 2 ed., John Wiley & Sons, 2016.

[8] J. C. Johnson, OCP: Oracle9i Performance Tuning Study Guide, John Wiley & Sons, 2006.

[9] TechTarget, "What is data warehouse?," February 2015. [Online]. Available: http://searchsqlserver.techtarget.com/definition/data-warehouse. [Accessed 14 August 2017].

[10] S. Dewan, Y. Aggarwal and S. Tanwar, "Review on Data Warehouse, Data Mining and OLAP Technology: As prerequisite aspect of business decision-making activity," IJRIT International Journal of Research in Information Technology, vol. 1, no. 10, pp. 30-39, October 2013.

[11] R. Gill and J. Singh, "A Review of Contemporary Data Quality Issues in Data Warehouse ETL Environment," Journal on Today's Ideas - Tomorrow's Technologies, vol. 2, no. 2, pp. 153-160, December 2014.

[12] I. A. T. Hashem, I. Yaqoob, N. B. Anuar, S. Mokhtar, A. Gani and S. U. Khan, "The rise of "big data" on cloud computing: Review and open research issues," Information Systems: Creation, Management and Utilization, pp. 98-115, 2015.

[13] F. Dehne, Q. Kong, A. Rau-Chaplin, H. Zaboli and R. Zhou, "Scalable real-time OLAP on cloud architectures," Journal of Parallel and Distributed Computing, 2014.

[14] J. George, V. Kumar and V. S. Kumar, "Data Warehouse Design Considerations for a Healthcare Business Intelligence System," Proceedings of the World Congress on Engineering, vol. 1, July 2015.

[15] R. Schiesser, IT System Management, Upper Saddle River, NJ: Prentice Hall PTR, 2010.

[16] M. T. a. R. Valverde, An Implementation of ITIL Guidelines for IT Support Process in a Service Organization, International Journal of Information and Electronics Engineering, Vol. 3, No. 3, May, 2013.

[17] A. S. Girsang and C. W. Prakoso, "Data Warehouse Development for Customer WIFI Access Service at a Telecommunication Company," *International Journal on Communications Antenna and Propagation (IRECAP),* vol. 7, no. 2, 2017.